Proceedings of the 10th Mediterranean Conference
on Control and Automation - MED2002
Lisbon, Portugal, July 9-12, 2002.

OPTIMISATION METHODS FOR IMPROVING FED-BATCH CULTIVATION OF *E. COLI* PRODUCING RECOMBINANT PROTEINS

I. Rocha, E.C. Ferreira

Centro de Engenharia Biológica, Universidade do Minho

4710–057 Braga PORTUGAL

e-mail: {irocha, ecferreira}@deb.uminho.pt

Keywords: Optimisation, fed-batch fermentation, *E. coli*, gradient methods, genetic algorithms.

Abstract

Two optimisation techniques for the fed-batch cultivation of high cell density *Escherichia coli* producing recombinant proteins were compared.

An unstructured model for the growth, based on the General State Space Dynamical Model [1] was used to represent the four major metabolic pathways: oxidative growth on glucose, fermentative growth on glucose, oxidative growth on acetate, and maintenance. The dilution rate (dependent on the substrate feed rate) was chosen as the input variable.

Recombinant protein production is known to be proportional, in our system, to the biomass concentration. Thus, biomass productivity was chosen as the criterion to be maximized.

The two methods compared were a first order gradient method based on Pontryagin's minimum principle and a stochastic method based on the biological principle of natural evolution, using a genetic algorithm. The former method revealed less efficient concerning to the computed maximum, and dependence on good initial values.

1 Introduction

Fermentation processes are commonly operated in fed-batch mode in order to prevent the accumulation of toxic substrates or products [10,23] thus allowing the achievement of higher product concentrations. The bacteria *Escherichia coli* is usually grown under this kind of operation due to the well-known negative effect of acetate, which is produced when the substrate, glucose, is presented above certain concentrations [7,22].

E. coli has been intensively studied in the last decades and has been widely used for the production of biopharmaceuticals [5,9]. In this kind of processes, the importance of operating near an optimal solution is evident due to the high value of the metabolites produced.

During a common fed-batch *E. coli* fermentation process, the system states change considerably, from a low initial cellular concentration to very high biomass and product concentrations. This dynamic behaviour motivates the development of optimisation methods to find the optimal input trajectories during the feeding stage in order to improve the process. The typical input variable in these processes is the substrate inflow rate as a function of time.

One way to evaluate the process performance is the measurement of the final product concentration. For the production of recombinant proteins in *E. coli* this is

usually related with the final biomass concentration obtained. Thus, a simple objective, although not the most accurate, is the maximisation of the final biomass concentration.

Many publications focus on this kind of optimisation problem in other bioprocesses [4,12-14,18,19,23], while only few [9,11] describe the application of optimisation algorithms in *E. coli* fermentation.

In this work two approaches were considered, namely the application of a gradient method and a genetic algorithm.

1 Process Modelling

Process simulation was conducted with a developed model derived from the general state space dynamical model described by Bastin and Dochain [1]. Generally, this approach can be described as representing the dynamics of *n* components and *m* reactions bioprocess:

$$\frac{d\xi}{dt} = K\varphi(\xi, t) - D\xi + F - Q \qquad (1)$$

where $\xi \in \mathbb{R}^n$ is a vector representing the state components; $K \in \mathbb{R}^{n\times m}$ is the yield coefficient matrix; $\varphi \in \mathbb{R}^m$ is the growth rates vector; the vectors F and Q are the feed rates and the gaseous outflow rates, respectively ($F, Q \in \mathbb{R}^n$). The scalar *D* is the dilution rate (D^{-1} is the residence time), which will be the manipulated variable, defined as follows:

$$D(t) = \frac{F_{in}(t)}{W(t)} \qquad (2)$$

being F_{in} the influent flow rate (kg h^{-1}), and *W* the culture medium weight (kg).

The growth of *E. coli* may be interpreted through four main reactions [3,6]:

Respiratory growth on glucose:

$$k_1 S + k_5 O \xrightarrow{m_1} X + k_7 C \qquad (3)$$

Fermentative growth on glucose:

$$k_2 S \xrightarrow{m_2} X + k_8 C + k_3 A \qquad (4)$$

Respiratory growth on acetate:

$$k_4 A + k_6 O \xrightarrow{m_3} X + k_9 C \qquad (5)$$

Maintenance:

$$S + k_{m1} A + k_{m2} O \xrightarrow{m_4} k_{m3} C \qquad (6)$$

where S, O, X, C, A represent sugar (glucose), oxygen, biomass, carbon dioxide, and acetate respectively (in the sequel *S*, *O*, *X*, *C*, and *A* mean concentrations); μ_1, μ_2, μ_3, and μ_4 are the growth rates; k_i are the yield (stoichiometric) coefficients.

The application of the general state space dynamical model to these reaction network leads to the following model:

$$\frac{d}{dt}\begin{bmatrix} X \\ S \\ A \\ O \\ C \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ -k_1 & -k_2 & 0 & -1 \\ 0 & k_3 & -k_4 & -k_{m1} \\ -k_5 & 0 & -k_6 & -k_{m2} \\ k_7 & k_8 & k_9 & k_{m3} \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix} X - D \begin{bmatrix} X \\ S \\ A \\ O \\ C \end{bmatrix} + \begin{bmatrix} 0 \\ DS_{in} \\ 0 \\ OTR \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ CTR \end{bmatrix} \qquad (7)$$

where *CTR* is the carbon dioxide transfer rate from liquid to gas phase, *OTR* is the oxygen transfer rate from gas to liquid phase, and S_{in} is the influent glucose concentration.

Depending on the feeding rate, *E. coli* growth can be taken simultaneously: i) from eq. 3 and eq. 6, ii) from eqs. 3, 4, and 6 or iii) from eqs. 3, 5, and 6. For the maximum biomass production, the preferred metabolic pathways correspond to the first case. Besides the lower biomass yield, the fermentative pathway (eq. 4) causes the production of acetate, an inhibitor of

biomass growth at relatively small concentrations (7.5 g/kg) [7,22]. Acetate consumption (eq. 5) yields a smaller biomass yield than glucose oxidative consumption (eq. 3).

In order to prevent the fermentation of glucose, it is necessary to keep the feeding at small values. On the other hand, those values must be sufficiently high in order to achieve the desired biomass concentration in a short time.

In the model, this behavior is governed by process kinetics, that is to say, from equations describing μ_1, μ_2, μ_3, and μ_4. These equations and parameters k_i used were based on the work described by [6].

3. Optimisation

Independently of the microorganism growth, the optimisation problem of a fed-batch fermentation is a typical problem of dynamic optimisation, consisting on finding a sequence of values for a control variable that minimize a chosen objective function, subject to the constraints represented by the dynamical model.

Thus, in this particular case, the problem consists on finding a sequence of values – the vector D that maximize the objective function:

$$J\left(t_f\right) = X\left(t_f\right) \qquad (8)$$

subject to the constraints represented on eq. 7.

Several optimisation methods have been applied to solve this kind of problem among them the two methods proposed in this paper, a first order gradient method based on Pontryagin's minimum principle and a genetic algorithm. Another popular method is the so-called Dynamic Programming. Although it is still not clear if this method has some advantages over other optimisation method in the particular case of fed-batch fermentations [18], it has been applied to the monoclonal antibody production optimisation [4] and to *Zymomous mobilis* fermentation [23]. This method was not applied to this particular case, as it requires the objective function to be a function of the states in every instant, which is not the case.

For both optimisation methods some parameters and constraints were based on a real *E. coli* fermentation system described elsewhere [16,17]. The final time was considered to be 25 hours, and the maximum feed rate was 0.4 kg/h.

3.1 Gradient Method

Gradient methods [2] based on Pontyagrin's minimum principle have been applied to the optimisation of generic bioprocesses [20] or hybridoma cells [18]. It is reported that, although a good final result are normally achieved, computation effort is considerable. Another aspect that has to be taken in consideration is the initial guess of the optimisation problem. According to [8], when the objective function is linear, which is the present case, the trajectories are very sensitive to the initial conditions. This author suggests, as a solution to this disadvantage, the definition of a quadratic cost criterion.

For the implementation of this method, the MATLAB 6.0 function '*fmincon*' was used. This function calculates the minimum of a function subject to several kinds of constraints starting at an initial estimate. It uses Sequential Programming methods, in which a Quadratic Programming (QP) subproblem is solved at each iteration [21]. An estimate of the Hessian of the Lagrangian is updated at each iteration using the BFGS formula. It requires as inputs the objective function, the vector of initial estimates, constraints, upper and lower bounds for the variables and a previously defined vector of options.

The objective function was evaluated in a script '*.m*' file, a routine that calculated the final biomass concentration as a function of the input vector D in an iterative way using 4th order Runge-Kutta to solve the differential equations of the model represented in eq. (7).

In order to optimise the computation time, the step of the integration was different from the length of the input D. Thus, vector D was divided in 25 elements (1 h^{-1})

that were interpolated 10 times for the calculation of the objective function value.

There were no constraint functions passed to the function *'fmincon'*, as the final weight restriction (5 kg) was accounted by a penalty in the objective function. The upper and lower bounds of the variable *D* were zero and the maximum feed rate, respectively.

The MATLAB routines were implemented in a Pentium III computer running at 933 MHz in a Windows 2000 environment.

Several simulations were conducted with the routines described. The results are presented in Table 1. The function converged to a solution, although it took 100 iterations to achieve a high final biomass concentration. It should also be mentioned that the values shown were obtained for an initial vector *D* linearly increasing with time from 0 to 0.1. With an initial constant vector *D* (for example of 0.05) the function did not converge to a high objective function, remaining in the values of 5 g/kg even after 1000 iterations. This is due to the already mentioned sensitivity of the method to initial values with linear objective functions.

After optimisation, the vector *D* corresponding to 1000 iterations was used to simulate a fermentation process during a 25 hours run. The results obtained are shown in figure 1.

Number of iterations	Objective function (g/kg)
Initial values	7.61
100	51.03
1000	53.91

Table 1. Results obtained with the gradient method after 100 and 1000 iterations.

3.2 Genetic Algorithm

Genetic Algorithms (GAs) have become a very attractive method of non-linear optimisation in recent years. GAs are based on natural evolutionary processes, where selection results in species that fit the best. Variation in the population is achieved by genetic cross-over and mutation.

These algorithms proved to be very suitable for the optimisation of highly non-linear problems with many variables. The application of these methods to fed-batch fermentations was done by several authors. Roubos *et al.* [19] and Nguang *et al.* [14] applied this approach to hybridoma cells, while yeast fermentation was the bioprocess studied by Na *et al.* [13].

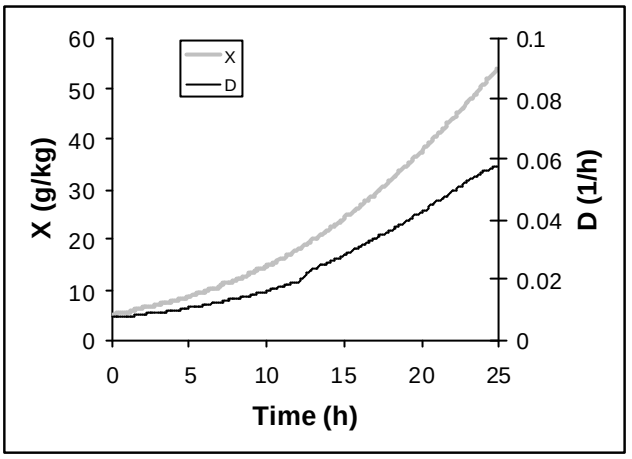

Figure 1. Evolution of the biomass and the input variable for the final solution of the gradient method (iteration 1000).

For the implementation of this method, a Genetic Algorithm Toolbox developed for MATLAB at the University of Sheffield was used [15]. It works with several genetic operators and supports binary, integer and real-valued representations. The representation chosen for this work was the real-valued, due to the advantages known [19]. The routine used as objective function was the same as in the previous method. The function *'tbxmpga'* was called from a high level script and conducted all the operations, namely, initialisation of the population, ranking and selection based on the objective function values, recombination/crossover, mutation, evaluation, reinsertion and migration. The number of individuals evaluated in any iteration was 200. Each individual was a sequence of values D_k, with values between upper and lower bounds mentioned above, composing a specific vector *D*.

The MATLAB routines were also ran in a Pentium III computer running at 933 MHz in a Windows 2000 environment.

The results obtained are illustrated in Table 2. It is clear that it took also 100 iterations to achieve a final solution slightly higher than the one obtained after 1000 iterations with the gradient method. The simulation of the fermentation with the vector *D* obtained gave the results shown in figure 2. The shape of the input vector as well as of the biomass concentration profile along the fermentation, are very similar to those observed for the previous method.

Number of iterations	Objective function (g/kg)
Initial values	5.51
50	41.72
100	53.93

Table 2. Results obtained with the GA method after 50 and 100 iterations.

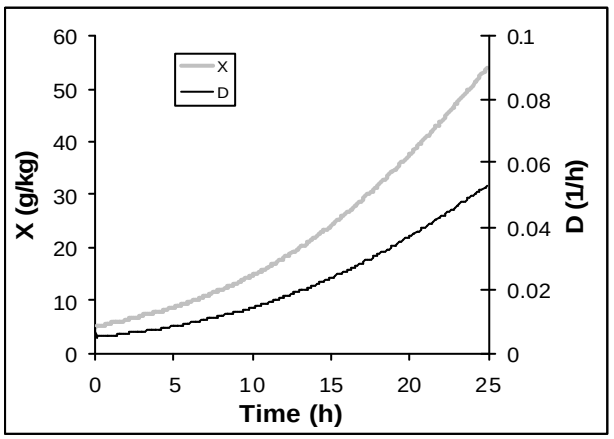

Figure 2. Evolution of the biomass and the input variable for the final solution of the GA (iteration 100).

4. Conclusions

The two methods under comparison revealed useful in achieving an optimal solution for the fed-batch fermentation of *E. coli*. However, the GAs revealed more efficient concerning to the computed maximum, and dependence on good initial values.

The simple objective function studied in this work should be changed into a quadratic criterion in future research in order to test the performance of the gradient method.

Acknowledgements

This work was supported by the PROTEXPRESS project (*Agência de Inovação*). *Fundação para a Ciência e a Tecnologia* provided financial support for I. Rocha through a doctoral research grant (PRAXIS XXI/BD/16961/98).

References

[1] G. Bastin, D. Dochain. *On-line Estimation and Adaptive Control of Bioreactors*. Elsevier Science Publishers, Amsterdam, (1990).

[2] A.E. Bryson, Jr. *Dynamic Opimization*, Addison-Wesley Longman, Inc., Menlo Park, 1999.

[3] A. Cockshott, I. Bogle. "Modelling the effect of glucose feeding on a recombinant E. coli fermentation", *Bioprocess Engineering*, **20**, pp. 83-90, (1999).

[4] M. de Tremblay, M. Perrier, C. Chavarie, J. Archambault, "Fed-Batch Culture of Hybridoma Cells: Comparison of Optimal Control Approach and Closed Loop Strategies", *Bioprocess Engineering*, **9**, pp. 13-21, (1993).

[5] J. Fieschko, T. Ritch, "Production of human alpha consensus interferon in recombinant Escherichia coli", *Chemical Engineering Communications*, **45**, pp. 229-240, (1986).

[6] V. Galvanauskas, R. Simutis, N. Volk, A. Lübert. "Model Based Design of a Biochemical cultivation process", *Bioprocess Engineering*, **18**, pp. 227-234, (1998).

[7] K. Han, H. C. Lim and J. Hong, "Acetic acid formation in Escherichia coli fermentation", *Biotechnology and Bioengineering*, **39**, pp. 663-671, (1992).

[8] A. Johnson, "Optimization of fed-batch biotechnical process", *IFAC Modeling and Control of Biotechnical Processes*, Pergamon Press, pp. 243-246, (1992).

[9] S.Y. Lee, "High cell-density culture of Escherichia coli", *Trends in Biotechnology*, **14**, pp. 98-105, (1996).

[10] J. Lee, S.Y. Lee, S. Park, A.P.J. Middelberg "Control of fed-batch fermentations", *Biotechnology Advances*, **17**, pp. 29-48, (1999).

[11] J. Lee, W.F. Ramirez, "Optimal fed-batch control of induced foreign protein production by recombinant bacteria" *A.I.Ch.E. J.* **40**, pp. 899-907, (1994).

[12] D. Leštan, A. Podgornik, A. Perdih, R. Stravs, "The sequential algorithm for dynamic optimization of bioprocesses", *IFAC Modeling and Control of Biotechnical Processes*, Pergamon Press, pp. 271-274, (1992).

[13] J.-G. Na, Y.K. Chang, B.H. Chung, H.C. Lim, "Adaptive optimization of fed-batch culture of yeast by using genetic algorithms", *Bioprocess and Biosystems Engineering*, **24**, pp. 299-308, (2002).

[14] S.K. Nguang, L.Z. Chen, X.D. Chen, "Optimisation of fed-batch culture of hybridoma cells using genetic algorithms", *ISA Transations*, **40**, pp. 381-389, (2001).

[15] H. Pohlheim, GEATbx: Genetic and Evolutionary Algorithm Toolbox for use with MATLAB – Documentation, (1995). www.systemtechnik.tu-ilmenau.de/~pohlheim/GA_Toolbox/index.html

[16] I. Rocha, E. C. Ferreira, "An Integrated System for Advanced Monitoring and Control of Fed-batch Fermentations of Recombinant *E. coli*", *8th International Conference on Computer Applications in Biotechnology* (D. Dochain, M. Perrier, Eds.), Québec City, Canada, pp 353-358, (2001).

[17] I. Rocha, E. C. Ferreira, "On-line simultaneous monitoring of glucose and acetate with flow-injection analysis during high-cell-density fermentation of recombinant *Escherichia coli"*, *Analytica Chimica Acta*, (accepted), (2002).

[18] J.A. Roubos, C.D. de Gooijer, G. van Straten, A.J.B. van Boxtel, "Comparison of optimization methods for fed-batch cultures of hybridoma cells", *Bioprocess Engineering*, **17**, pp. 99-102, (1997).

[19] J.A. Roubos, G. van Straten, A.J.B. van Boxtel "An evolutionary strategy for fed-batch bioreactor optimization; concepts and performance", *Journal of Biotechnology*, **67**, 173-187, (1999).

[20] S. Shioya, "Optimization and control for maximum production in fed-batch culture", *IFAC Modeling and Control of Biotechnical Processes*, Pergamon Press, pp. 117-122, (1992).

[21] The Mathworks Inc. *Optimization Toolbox User's Guide*. The Mathworks Inc, Natick, MA, (2000).

[22] M. van de Walle, J. Shiloach, "Proposed mechanism of acetate accumulation in two recombinant Escherichia coli strains during high density fermentation", *Biotechnology and Bioengineering*, **57**, pp. 71-78, (1998).

[23] F.S. Wang, C.H. Shyu, "Optimal feed policy for fed-batch fermentation of ethanol production by Zymomous mobilis", *Bioprocess Engineering*, **17**, pp. 63-68, (1997).

[24] T. Yamane, S. Shimizu, "Fed-batch Techniques in Microbial Processes", *Advances in Biochemical Engineering*, **30**, pp. 147-194, (1984).